OPTICAL CLUSTER FINDING WITH AN ADAPTIVE MATCHED-FILTER TECHNIQUE: ALGORITHM AND COMPARISON WITH SIMULATIONS

Feng Dong,[1] Elena Pierpaoli,[2] James E. Gunn,[3] and Risa H. Wechsler[4]

ABSTRACT

We present a modified adaptive matched-filter algorithm designed to identify clusters of galaxies in wide-field imaging surveys such as the Sloan Digital Sky Survey (SDSS). The cluster-finding technique is fully adaptive to imaging surveys with spectroscopic coverage, multicolor photometric redshifts, no redshift information at all, and any combination of these within one survey. It works with high efficiency in multiband imaging surveys for which photometric redshifts can be estimated with well-understood error distributions. Tests of the algorithm on realistic mock SDSS catalogs suggest that the detected sample is ~85% complete and over 90% pure for clusters with masses above $1.0 \times 10^{14} \, h^{-1} \, M_{\odot}$ and redshifts up to $z = 0.45$. The errors of estimated cluster redshifts from a maximum likelihood method are shown to be small (typically less than 0.01) over the whole redshift range, with photometric redshift errors typical of those found in the SDSS. Inside the spherical radius corresponding to a galaxy overdensity of $\Delta = 200$, we find the derived cluster richness Λ_{200} to be a roughly linear indicator of its virial mass M_{200}, which well recovers the relation between total luminosity and cluster mass of the input simulation.

Subject headings: cosmology: theory — galaxies: clusters: general — large-scale structure of universe

1. INTRODUCTION

Clusters of galaxies are the most massive virialized systems in the universe, and have been extensively used to study galaxy population and evolution (Dressler 1984; Dressler & Gunn 1992), to trace the large-scale structure of the universe (Bahcall 1988; Postman et al. 1992), and to constrain cosmology (Evrard 1989; Bahcall et al. 1999; Henry 2000; Pierpaoli et al. 2001, 2003). Given the important role that clusters of galaxies play in the studies of both astrophysics and cosmology, tremendous efforts have been made during the past several decades to search for these systems. The first large samples of clusters were identified by looking for projected galaxy overdensities through visual inspection of photographic plates (Abell 1958; Abell et al. 1989; Zwicky et al. 1968). These catalogs made pioneering contributions to our understanding of the extragalactic universe, and since their generation have opened many new frontiers in the studies of galaxy clusters. However, the compilation of a relatively complete and pure sample of galaxy clusters has remained far from trivial. To date, the Abell catalog, which contains about 4000 rich clusters to a redshift of $z \sim 0.2$, is still the most widely used cluster catalog in the field, although it was realized early that visually constructed catalogs suffer from projection effects, subjectivity, and large uncertainties in estimated properties (Sutherland 1988). It is difficult to use these catalogs for statistical studies in cosmology because of these uncertainties, and because the selection function and false-positive rates of such cluster samples are hard to quantify.

To relieve some of these concerns, other approaches for identifying clusters have also been designed and implemented, such as the reconstruction of the full three-dimensional structures in complete redshift surveys (Huchra & Geller 1982; Geller & Huchra 1983; Ramella et al. 1997), the detection of clusters in X-ray surveys (Gioia et al. 1990; Edge et al. 1990; Ebeling et al. 1998; Rosati et al. 1998; Romer et al. 2000; Scharf et al. 2000; Böhringer et al. 2001, 2004; Mullis et al. 2003), and the utilization of the Sunyaev-Zeldovich effect (Carlstrom et al. 2000; Mohr et al. 2002; Pierpaoli et al. 2005) and weak gravitational lensing (Schneider 1996; Wittman et al. 2001) in the search for clusters. Moreover, the realization of large and deep galaxy surveys in recent years has revived optical cluster-finding endeavors, and prompted the development of more automated and rigorous algorithms to select clusters from imaging surveys. Using multicolor photometric data from which photometric redshifts can be estimated, it is now possible to mitigate the problems of projection effects, and to quantitatively analyze the selection bias. Automated peak-finding techniques in optical cluster searches were attempted by Shectman (1985), and later used in the Edinburgh/Durham survey (Lumsden et al. 1992), as well as the Automatic Plate Measurement Facility survey (Dalton et al. 1994; Dalton et al. 1997). In the construction of the cluster catalog from the Palomar Distant Cluster Survey (Postman et al. 1996), a matched-filter algorithm was developed to select clusters from a photometric galaxy sample. It was widely used in subsequent surveys, and several variants have been put forward (Kawasaki et al. 1998; Schuecker & Boehringer 1998; Kepner et al. 1999; Kim et al. 2002; White & Kochanek 2002). Meanwhile, with knowledge of the existence of the E/S0 ridgeline of cluster galaxies in color-magnitude space and the aid of multicolor CCD photometry, several color-based cluster-finding techniques have also been investigated (Gladders & Yee 2000, 2005; Goto et al. 2002; Miller et al. 2005). Some of these have already been successfully applied to select clusters from the Sloan Digital Sky Survey (SDSS) data (Goto et al. 2002; Annis et al. 2002; Bahcall et al. 2003; Miller et al. 2005; Koester et al. 2007).

The SDSS (York et al. 2000) is a five-band CCD imaging survey of about $10^4 \, \mathrm{deg}^2$ in the high-latitude north Galactic cap and a smaller deeper region in the south, followed by an extensive multifiber spectroscopic survey. The imaging survey is carried out in drift-scan mode in five SDSS filters (u, g, r, i, and z) to a limiting magnitude of $r \sim 22.5$ (Fukugita et al. 1996; Gunn et al. 1998; Lupton et al. 2001; Smith et al. 2002). The spectroscopic

[1] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544; feng@astro.princeton.edu.
[2] University of Southern California, Los Angeles, CA 90089-0484; pierpaol@usc.edu.
[3] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544; jeg@astro.princeton.edu.
[4] Kavli Institute for Particle Astrophysics and Cosmology, Physics Department, and Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94305.

survey targets $\sim 10^6$ galaxies to $r \sim 17.7$, with a median redshift of $z \sim 0.1$ (Strauss et al. 2002), and a smaller deeper sample of $\sim 10^5$ luminous red galaxies (LRGs) out to $z \sim 0.5$ (Eisenstein et al. 2001). In this paper, we discuss a modified adaptive matched-filter technique that incorporates several new features absent in previous algorithms, and which is designed to detect clusters using both the SDSS imaging and spectroscopic data; it could readily be adapted to other similar multiband, large-area galaxy surveys for construction of optically selected cluster samples. This is the first of a series of papers that will explore the application of the technique to select clusters from the SDSS.

The general idea of the matched-filter method relies on the fact that clusters show, on average, a typical density profile, now widely assumed to be the NFW form first suggested by Navarro, Frenk, and White (Navarro et al. 1996). Assuming that galaxies trace dark matter, we expect galaxies within clusters to be distributed according to such a profile. The algorithm selects regions in the sky where the distribution of galaxies corresponds to the projection of the average cluster density profile. In addition, it is possible to specify galaxy redshift information inside clusters, and to use prior knowledge of the galaxy luminosity function. The combination of these matched subfilters thus enables us to extract a quantitative signal corresponding to the existence of a cluster at a given location in the surveyed sky area.

The modified matched-filter technique presented in this paper can fully adapt to imaging surveys with spectroscopic measurements, multicolor photometric redshifts, no redshift information at all, and any combination of these within one survey. In the SDSS, where photometric redshifts can be estimated with well-understood error distributions from the five-band (u, g, r, i, and z) multicolor photometry, the matched-filter technique described here utilizes not only the spectroscopic coverage for the bright main sample galaxies and LRGs, but also the photometric redshift information for most of the galaxies detected in the imaging survey. This expands the input galaxy sample that can be fed into the cluster-finding algorithm far beyond what is obtainable by pure spectroscopic methods (e.g., Miller et al. 2005). The obtained composite cluster catalog can also go much deeper in redshift ($z \sim 0.4$–0.5 in this case) than the typical $z \sim 0.2$ limit for spectroscopic samples, due to the lack of availability of spectroscopic measurements for faint, deep galaxies.

Since the matched-filter technique does not explicitly use the information about the red sequence to select clusters, as is done in some color-based cluster-finding methods (Annis et al. 2002; Miller et al. 2005; Koester et al. 2007), it can theoretically detect clusters of any type in color, and is not restricted only to old, red E/S0 galaxies. Such clusters likely dominate the cluster population, but may not constitute all of it, especially as one probes systems of lower richness and at higher redshifts. The use of both spectroscopic and photometric redshift information largely eliminates projection effects and removes most of the phantom clusters. The matched filter also generates accurate quantitative estimates of derived cluster properties such as redshift, scale, richness, and concentration, and produces quantitative detection likelihoods, indicative of the combined information for both red and blue galaxies identified as cluster members. These facilitate further studies of detected systems and make easier the comparison to clusters selected by other methods. One major concern for the matched-filter technique is the fact that determination of these parameters depends on the specific cluster model we use to build the relevant filters. However, these effects can be minimized by careful assumptions about the shape and evolution of the luminosity function, and by the fact that our density filter is self-adaptive to different cluster scales and concentrations. The clusters selected by the al-

gorithm will provide the necessary sample on which we then apply an iterative procedure aimed at refining the constraints on the clusters' properties. More details will be discussed in § 2 and subsequent work following this paper.

The new algorithm presented here differs from previous matched-filter implementations (Kepner et al. 1999; Kim et al. 2002) in several ways. We use a uniform Poisson likelihood analysis, which is only the second step in the approach by Kepner et al. (1999), following a first pass using Gaussian statistics for preselection of clusters. This avoids the common problem for high-redshift clusters of having too few galaxies in any cell of interest for Gaussian statistics to apply; our approach yields correct likelihoods even at the detection stage. In addition, both the core radius and virial radius of the matched filter are adaptive over the typical observed dynamical range for clusters, in contrast to most previous cluster-finding techniques, which set the cluster core radius or search radius at a fixed value. For each individual cluster, a best-fit core radius is found to maximize the likelihood match, as well as an outer radius inside which the galaxy overdensity reaches $\Delta = 200$. The cluster richness is then normalized to be the light contained within this virial radius, which we find correlates better with the mass of gravitational systems whose extent is defined by density contrast, as is widely adopted in theoretical studies. The new features of our modified algorithm will be further discussed in § 2.

In order to understand the biases and the selection functions of our algorithm, we test it on a mock SDSS catalog, which has been constructed from the Hubble Volume Simulation (Evrard et al. 2002) by assigning luminosities and colors to the dark matter particles in a manner that reproduces many characteristics of the galaxy population from SDSS observations. The "observations" of the simulations are then further modified so that the redshift scatter of those galaxies that have photometric, but no spectroscopic, redshifts corresponds to that of the photometric redshift errors in actual SDSS data. The comparison of the detected cluster sample with halos in the simulation provides the only rigorous way to assess how the observed cluster properties relate to the real masses, and how the cluster sample can be used to derive cosmological constraints.

In § 2, we describe the modified adaptive matched-filter technique, and how it is used to extract the cluster sample. Section 3 presents the basic features of the simulated catalog that we adopted for testing purposes. In § 4, we show results on the completeness and purity of our cluster sample, and the expected scaling relations inferred from runs on the simulations. We conclude in § 5.

A flat ΛCDM model with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ is used throughout this work, and we assume a Hubble constant of $H_0 = 100\,h$ km s^{-1} Mpc^{-1} if not specified otherwise.

2. THE CLUSTER-FINDING ALGORITHM

The matched-filter technique introduced here is a likelihood method which identifies clusters by convolving the optical galaxy survey with a set of filters based on a modeling of the cluster and field galaxy distributions. A cluster radial surface density profile, a galaxy luminosity function, and redshift information (when available) are used to construct filters in position, magnitude, and redshift space, from which a cluster likelihood map is generated. The peaks in the map thus correspond to candidate cluster centers, where the matches between the survey data and the cluster filters are optimized. The algorithm automatically provides the probability for the detection, best-fit estimates of cluster properties including redshift, radius, and richness, as well as membership assessment for each galaxy. The modified algorithm can be fully adapted to current and future galaxy surveys in 2D (imaging), 2.5D (in which multicolor photometric redshifts and their errors

can be estimated), and 3D (with full spectroscopic redshift measurements). Using the apparent magnitudes and, where applicable, the redshift estimates, instead of simply searching for projected galaxy overdensities effectively suppresses the foreground-background contamination, and the technique has proven to be an efficient way of selecting clusters of galaxies from large multiband optical surveys.

In what follows, we first provide a general introduction on how the likelihood function is constructed, and how we detect clusters with the matched-filter method. This gives us an overview of how the cluster catalog is derived. Then we discuss in more detail the density models and subfilters used to construct the likelihood. More specifically, we assume an NFW density profile, a general Schechter luminosity function, and a Gaussian model for BCGs to model clusters, and we use the spectroscopic measurements and obtained error distributions of galaxy photometric redshifts from the SDSS to incorporate redshift uncertainties. Finally, we describe how to determine the set of best-fit parameters on cluster properties that maximizes the likelihood at a given position over a range of redshifts, scales, concentrations, and richnesses.

2.1. Likelihood Function

The likelihood function used here is based on the assumption that the probability of finding galaxies in an infinitesimal bin in angular position, apparent magnitude, and redshift space is given by a Poisson distribution. Under this assumption, the total likelihood of many of such bins, which we take to be centered in the location of the galaxies in the survey, is

$$\ln \mathcal{L} = -N_f - \sum_{k=1}^{N_c} N_k + \sum_{i=1}^{N_g} \ln P(i), \qquad (1)$$

where N_f is the total number of field galaxies expected within the searching area, N_g is the total number of galaxies, and $\sum_{k=1}^{N_c} N_k$ is normalized to be the number of galaxies brighter than L^* as members of the N_c clusters assumed in the model (see Appendix C2 in Kepner et al. 1999 for a full derivation). Here, $P(i)$ represents the predicted probability density of galaxies in a given bin, which includes both probabilities of field galaxies (P_f) and of cluster members (P_c):

$$P(i) = P_f(i) + \sum_{k=1}^{N_c} P_c(i, k). \qquad (2)$$

These probabilities are the expected number densities for a given location and magnitude.

The cluster catalog is constructed with an iterative procedure similar to the one used in Kochanek et al. (2003). We start our process from a density model of a smooth background with no clusters. For each galaxy position, we then evaluate the likelihood increment we would obtain by assuming that there is in fact a cluster centered on that galaxy. The likelihood is then optimized by varying the cluster galaxy number N_k, the redshift, and the cluster scale length. At each iteration, we retain the cluster candidate that results in the greatest likelihood increase. We incorporate it into our density model, and restart the procedure. Therefore, the function for finding the kth cluster in the whole surveyed area is

$$\Delta \ln \mathcal{L}(k) = -N_k + \sum_{i=1}^{N_g} \ln \left[\frac{P_f(i) + \sum_{j=1}^{k} P_c(i, j)}{P_f(i) + \sum_{j=1}^{k-1} P_c(i, j)} \right]. \qquad (3)$$

A list of cluster candidates then becomes available in decreasing order of detection likelihoods. For each candidate, one has derived

properties, including best-fit position, scale, richness, and estimated redshift. The initial cluster catalog allows us to further inspect each individual candidate for exploration of substructure and place better constraints on previously fitted quantities.

2.2. Density Model

As both field and cluster galaxies are found in the survey, the probability of finding a galaxy in a given bin depends on the density of both of these populations (see eq. [2]). For a galaxy i with angular position $\boldsymbol{\theta}_i$, r-band apparent magnitude m_i^r, and redshift z_i (when available), the background number density $P_f(i)$ can be directly extracted from the global number counts of the galaxy survey,

$$P_f(i) = \frac{dN}{dm\, dz}(m_i^r, z_i), \qquad (4)$$

although this must be modified to account for the effects of galaxy redshift uncertainties if photometric redshift estimates are used.

For a cluster k located at $\boldsymbol{\theta}_k$ with proper scale length r_{ck}, redshift z_k, and galaxy number N_k, the probability of galaxy i being a member of k, $P_c(i, k)$, is just the product of a surface density profile Σ_c and a luminosity function ϕ_c at the cluster's redshift, times a distribution function $f(z_i - z_k)$ that expresses redshift uncertainties:

$$P_c(i, k) = N_k \Sigma_c[D_A(z_k)\boldsymbol{\theta}_{ik}] \phi_c [m_i^r - \mathcal{D}(z_k)] f(z_i - z_k), \qquad (5)$$

where $\mathcal{D}(z_k)$ is defined through

$$M_i^r = m_i^r - 5 \log[D_L(z_k)/10 \text{ pc}] - k(z_k) = m_i^r - \mathcal{D}(z_k), \qquad (6)$$

where $D_A(z_k)$ and $D_L(z_k)$ are the angular diameter and luminosity distance, respectively, at the cluster's redshift z_k, and where $k(z_k)$ is the k-correction. The conversion of units in luminosity and distance is conducted by performing proper k-corrections for galaxies of different spectral types and by choosing the proper cosmology (see § 1).

2.3. Subfilters

Based on current observational studies and findings from dark matter halos, as well as for convenient comparisons to theoretical models widely used in analytical studies and N-body simulations, we assume that the density profile of galaxies within a cluster follows the form of a NFW profile (Navarro et al. 1996), which in three dimensions is given by

$$\rho_c(r) = \frac{1}{4\pi r_c^3 F(c)} \frac{1}{(r/r_c)(1 + r/r_c)^2}, \qquad (7)$$

where c is the concentration parameter and $F(c)$ is the typical normalization factor for galaxies inside the virial radius of the cluster, $r_v = c r_c$. The 3D profile is then integrated along the line of sight to derive a projected surface density profile $\Sigma_c(r)$, which is expressible as a much more complicated analytical form (see Bartelmann 1996). The profile is normalized so that $\int_0^{c r_c} 2\pi r \Sigma_c(r)\, dr = 1$.

The search radius for galaxies belonging to the cluster is set to be the virial radius of the cluster, or more specifically here, the radius inside which the mass overdensity is 200 times the critical density, i.e., $200\Omega_m^{-1}$ times the average background (Evrard et al.

2002). Since it is hard to directly measure the cluster mass over-density in observations, we instead determine the virial radius inside which the space density of cluster galaxies is $200\Omega_m^{-1}$ times the mean field, assuming that the galaxy distribution in a halo traces the overall dark matter distribution (see discussions in Hansen et al. 2005), which has been suggested by recent observations and simulations (Lin et al. 2004; Nagai & Kravtsov 2005; Lin & Mohr 2007), and is supported by weak-lensing measurements (Sheldon et al. 2004). For simplicity, we use r_{200} throughout this work to denote the cluster virial radius determined by galaxy overdensities. The cluster richness is then defined to be the total luminosity in units of L^* inside r_{200}.

As has been discussed before in matched-filter studies (Postman et al. 1996; Kim et al. 2002) and also shown by our own numerical experiments, the efficiency of the filter is usually much more sensitive to the overall filter cutoff radius than to the details of its shape. Therefore, the determination of appropriate values for the scale length in the cluster model is of particular importance, as it may have a significant impact on the detection efficiency of the cluster-finding algorithm. Most previous matched-filter methods have used a carefully chosen fixed value for the model cluster cut-off radii, and they compute the galaxy number or the richness of clusters within such a fixed radius in physical units. Postman et al. (1996) conclude that a fixed search radius of 1 Mpc h^{-1} is a near-optimal choice in their radial filter, and this value has also been adopted by Kepner et al. (1999) and Kim et al. (2002) in their method, which assumes a modified Plummer-law model for the surface density profile. In White & Kochanek (2002) and Kochanek et al. (2003), the authors set a fixed core radius of $r_c = 200$ kpc h^{-1} and a concentration parameter of $c = 4$ for the NFW profile in the cluster detection and mass estimates. Although we find from observations and simulations that these choices are reasonable values for typical rich clusters, a single fixed scale length for all clusters over a wide range of masses and concentrations will certainly degrade the signal-to-noise ratio and bias detection probabilities, and will be responsible for at least part of the large scatter observed in previous cluster mass-richness scaling relations. In our modified adaptive matched-filter algorithm, we optimize the core radius for each individual cluster over the dynamical range for typical galaxy clusters. For the core radius value that maximizes the likelihood, we then compute the normalized cluster richness according to the NFW profile with best-fit parameters within a cluster virial radius r_{200}, determined from galaxy overdensities. We believe that this procedure is more similar to and comparable with the virial mass as defined by density contrast, which is used in most theoretical studies and analyses of simulations.

For the magnitude filter, we adopt a luminosity profile described by a central galaxy plus a standard Schechter luminosity function (Schechter 1976)

$$\phi(M) = \frac{dn}{dM} = 0.4 \ln 10 n^* \left(\frac{L}{L^*}\right)^{1+\alpha} \exp\left(\frac{-L}{L^*}\right); \quad (8)$$

and the integrated luminosity function is

$$\Phi(M) = \int_{-\infty}^{M} \phi(M)\, dM = n^* \Gamma(1 + \alpha, L/L^*). \quad (9)$$

Parameters for the global luminosity function are obtained from the SDSS spectroscopic sample at the redshift of $z = 0.1$ (Blanton et al. 2003). To account for the evolutionary effects at higher red-shifts, we allow a passive evolution of L^*, which brightens about 0.8 mag from $z = 0$ to $z = 0.5$ (Loveday et al. 1992; Lilly et al. 1995b; Nagamine et al. 2001; Blanton et al. 2003; Loveday 2004; Baldry et al. 2005; Ilbert et al. 2005). We assume that L^* does not vary as a function of cluster richness, an assumption supported by the results of Hansen et al. (2005). Because the matched-filter algorithm uses both a cluster galaxy luminosity function and a field galaxy luminosity function, which are expected to be different due to the morphology-density relation (Dressler 1980) and the observed dependence of the luminosity function on galaxy over-densities (Christlein 2000; Mo et al. 2004; Croton et al. 2005), it would be desirable to model these separately. It would also be desirable to further model the luminosity distributions according to galaxy spectral types (Folkes et al. 1999; Lin et al. 1999; Hogg et al. 2003). At this stage, however, only a single function has been adopted, since the work on precise luminosity functions for cluster galaxies of different types has just been started. We hope to investigate this further on the basis of the first catalog we produce. Once a cluster catalog is available for galaxies in all red-shift ranges, we can examine the impact of our assumptions about the galaxy luminosity functions, as well as their evolution for different environments and spectral types. In order to use the same range in the luminosity function at all distances, and therefore avoid the bias associated with errors in the assumed form of the luminosity function, we cut off the luminosity function at 1 mag below L^*. We can still calculate total luminosities by integrating the assumed form, and we use this in our richness calculation, described below.

The existence of brightest cluster galaxies (BCGs) near the cluster centers is incorporated into our cluster galaxy luminosity model as a component separate from the main Schechter function for satellites, as this distinction has been clearly seen in clusters over a range of richnesses (Tremaine & Richstone 1977; Hansen et al. 2005). We assume a Gaussian distribution for the luminosities of these objects, and adopt the results from Lin & Mohr (2004) for correlations between the BCG luminosity and host cluster properties. More specifically, the BCG luminosity is assumed to follow a single power law with a cluster richness $L_{\rm BCG} \sim \Lambda_{200}^{1/4}$, and we take the width of the Gaussian to be $\sim$0.5 mag (Lin & Mohr 2004; Zheng et al. 2005; Hansen et al. 2005). The luminosity of BCGs is assumed to evolve in the same way as L^* does, i.e., the luminosity at the mean of the Gaussian has a constant ratio to L^*. This is almost certainly incorrect in detail, but will be explored in follow-up work once the catalog is constructed. This modification of the general Schechter function enhances the detectability of typical clusters with BCGs, especially those at higher redshifts with only few galaxies other than the BCG to be included in the apparent magnitude-limited galaxy sample.

Thanks to the accurate five-band (u, g, r, i, and z) multicolor photometry in the SDSS (York et al. 2000), as well as the associated redshift survey for the bright main sample galaxies (Strauss et al. 2002) and LRGs (Eisenstein et al. 2001), it is now also possible to retrieve redshift information for most of the galaxies that we are going to use in the construction of the SDSS cluster catalog, either photometrically or spectroscopically. For real SDSS data currently available from DR5, we find that galaxies with valid photometric redshift estimates make up more than 96% of the whole sample in the imaging data, within which about 1%, mostly bright, red galaxies, have matched spectroscopic measurements from red-shift surveys. Not surprisingly, the inclusion of galaxy redshift estimates greatly improves the accuracy of the cluster redshift determinations and significantly mitigates projection effects, thus allowing the detection of much poorer systems than was possible in previous work with no redshift measurements.

The uncertainties of galaxy redshifts are assumed to follow Gaussian distributions in the 2.5D and 3D cases, where in terms of the $f(z)$ function in equation (5) we have

$$f(z_k) = \frac{\exp\left[-(z_i - z_k)^2/2\sigma_{z_i}^2\right]}{\sqrt{2\pi}\sigma_{z_i}}. \tag{10}$$

For galaxies with computed photometric redshifts (described below), we add to the cluster galaxy density model a third sub-filter based on the distribution of derived redshift uncertainties in the form of a combination of multiple Gaussian modes. These error estimates are obtained by calibrating photometric red-shifts with the real redshifts in the SDSS spectroscopic galaxy sample and redshifts for other fainter (but smaller) overlapping surveys. The analysis is done for red and blue galaxies separately using the color separator by Strateva et al. (2001), and it is found that a model using Gaussian modes with proper weights generally provides a good description of the bias and scatter in the photometric redshifts for galaxies of both spectral types and in different apparent magnitude bins. Some of the results are shown in § 3.

In the 3D case in which spectroscopic redshifts of galaxies are measured, we smooth the results in Gaussians with assigned cluster velocity dispersions that vary in a range from 400 km s^{-1} (proper) for poorer clusters to 1200 km s^{-1} (proper) for the richest systems in the selected cluster sample, according to several discrete estimated richness classes. The same procedure outlined in the previous paragraph for photometric redshifts is applied to include this redshift filter in the galaxy density model.

In addition, we find galaxies that either have invalid photometric redshifts computed or that fall into a redshift and magnitude range where no good calibrations are available. Such galaxies, which are currently about 3%–5% of the whole sample, are assumed to have no redshift estimates, and therefore no constraining filter. Hence, we set up the appropriate scenario for each galaxy that adapts the matched-filter algorithm to galaxy redshift estimates with varied accuracy.

Finally, of course, we fit an overall amplitude, which represents the cluster richness. Since its size, shape, and redshift are all determined at this point, we can express the amplitude however we like in physical terms. We have chosen to use the total luminosity within r_{200} expressed as a multiple of L^* (evolved to the relevant redshift using 1.6 mag of luminosity evolution per unit redshift), which we denote as Λ_{200}.

2.4. Implementation

Implementation of the matched-filter algorithm starts with reading the galaxy catalog. For each galaxy i in the sample, we read in the positions α_i and δ_i, the extinction-corrected five-band apparent magnitudes and their errors, and the redshift z_i if it has a matched spectrum. Using the flux and color information, we compute a photometric redshift estimate using a neural network technique by Lin et al. (2006), as well as k-corrections and estimated rest-frame colors for each galaxy, which we add as input to the cluster-finding algorithm.

The next step is to define the cluster model we adopt for the filters, including the surface density profile $\Sigma_c(r)$, the luminosity function $\phi(M)$, and the assumed Gaussian modes of photometric redshift uncertainties. The field density model $P_f(m, z)$ is constructed from global number counts of the surveyed background galaxy distributions as a function of magnitude and redshift, as shown in equation (4). We then incorporate these models into the Poisson likelihood functions as discussed above.

To map the likelihood distributions of the surveyed area, we grid the sky using the Healpix package of Górski et al. (2005), which provides a useful hierarchical pixelization scheme of equal-area pixels. In Kepner et al. (1999), the authors calculate the like-lihoods by choosing galaxy positions on an adaptive grid, instead of the using the uniform grids used in previous matched-filter codes (Postman et al. 1996); this ensures sufficient resolution in the high-density regions, while saving computational time and memory for less dense regions. We follow this procedure, and evaluate the likelihood functions at every galaxy position to locate the peaks in the map as possible cluster centers. The cluster richness is optimized over the whole redshift range of our search at intervals that converge to $\delta z = 0.001$, and for a set of trial cluster scale radii (r_c) at 10 kpc h^{-1} steps. The derived quantities for best-fit cluster richness, redshift, and scale length thus correspond to the parameters that maximize the likelihood function at the grid position or candidate cluster center.

This algorithm possesses several new features. First, the cluster algorithm is fully adaptive to 2D, 2.5D, and 3D cases in the optical surveys, and can deal with data with these different attributes simultaneously. It can easily accommodate galaxy redshifts with uncertainties in any forms and distributions, from purely single-band imaging data to a complete spectroscopic redshift survey, and it works well for the intermediate case in which photometric redshifts are estimated from multiband color information. Projection effects from foreground-background contamination, which have been a long-standing problem for optically selected clusters, are largely suppressed. This allows the detection of even poorer systems at high redshift, and shows great potential for current and future large, deeper surveys in the optical band. Second, the current adaptive matched filter used single-Poisson statistics in the likelihood analysis, compared to the two-step approach in Kepner et al. (1999), which uses a coarse filter based on Gaussian likelihood for preselection of clusters. We wrote our code in Fortran 90, and by careful arrangement in computations and in setting up the quick link search, the optimization of the Poisson likelihood through the whole process is now affordable in terms of execution time and memory. For a survey field of ~300 deg^2, which is comparable to a typical SDSS stripe (York et al. 2000), the modified adaptive matched-filter algorithm requires around 900 Mbytes of memory and takes about 30 hr for a single run using one dual-processor node in a Linux Beowulf cluster with a clock speed of 3.06 GHz per processor. With no necessity that there be a minimum number of galaxies inside each virtual bin, as is the case in the Gaussian algorithm, the Poisson statistics remains robust in the common situation where there are too few galaxies in each cell for Gaussian statistics to apply. Third, as discussed in White & Kochanek (2002) and Kochanek et al. (2003), the current density model explicitly includes the effect of previously found clusters on the global likelihood function. The procedure automatically separates overlapping clusters and avoids multiple detections of the same system in the overdensity regions, and is thus somewhat similar to the CLEAN method used in radio astronomy to produce maps (Högbom 1974; Schwarz 1978). We do not need to do extra cluster deblending work afterward. Finally, as discussed earlier, our approach to maximizing the likelihood differs from most previous cluster-finding techniques that choose a fixed cluster scale or search radius. We optimize the core radius for each individual cluster, and the cluster richness is computed within a virial radius determined from galaxy overdensities. This provides insights about the virial mass of such
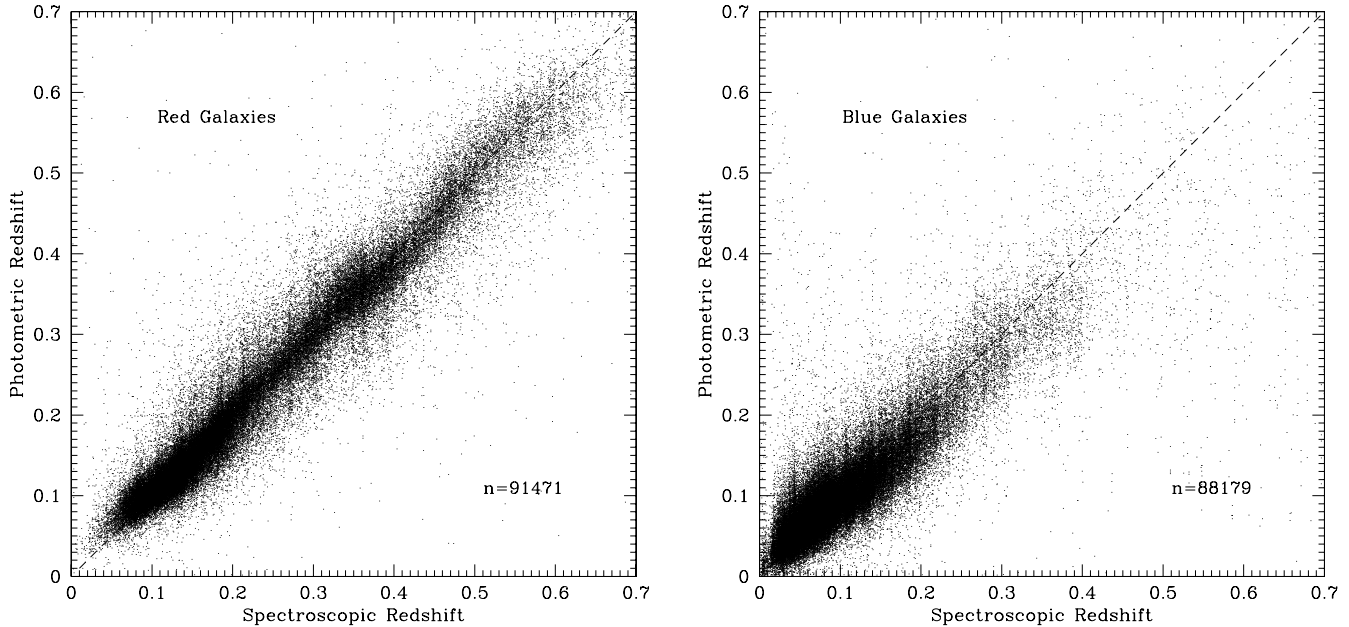
Fig. 1.—Calculated photometric redshifts vs. corresponding spectroscopic measurements for early-type galaxies (or red galaxies, *left*), and late type galaxies (or blue galaxies, *right*). Here, red means $g - r > 1.3$, and blue means $g - r < 1.3$.

gravitational systems defined by density contrast, and better corresponds to what is done in theoretical treatments.

3. TESTS ON MOCK GALAXY CATALOGS

To evaluate the completeness and purity (false-positive rate) of our cluster sample, as well as to assess how well our measured cluster properties correspond to the properties of the underlying dark matter halos, we run the matched-filter algorithm on a mock galaxy catalog generated from a realistic cosmological N-body simulation. Because of the large redshift range we are trying to probe, it is important to do this with as large a simulation volume as possible. In addition, because we seek here to test the behavior of our algorithm using a combination of spectroscopic and photometric redshifts, it is useful to have a realistic galaxy population in both the clusters and the field, with luminosities, colors, and a relation between these quantities and environment that are a good match to SDSS data. Here we have used a mock catalog based on ADDGALS (adding density-determined galaxies to lightcone simulations; Wechsler 2004; R. H. Wechsler 2008, in preparation), a method designed to model relatively bright galaxies in large-volume simulations.

The underlying dark matter simulation used here tracks 10^9 particles of mass $2.25 \times 10^{12} \, h^{-1} \, M_\odot$ in a periodic cubic volume with side length of 3 h^{-1} Gpc, using a flat ΛCDM cosmology with $\Omega_m = 0.3$, $\sigma_8 = 0.9$, and $h = 0.7$ (the Hubble Volume Simulation; Evrard et al. 2002). Halos are identified for masses above $2.7 \times 10^{13} \, h^{-1} \, M_\odot$. Data are collected on the past light cone of an observer at the center of the volume. The size of the simulation enables the creation of a full-sky survey out to redshift of $z = 0.58$, and is thus suited to testing our cluster-finding algorithm out to high redshifts using the SDSS imaging data.

Galaxies are connected to individual dark matter particles on this simulated light cone, and are subject to several empirical constraints. The resolution of the simulation allows the mock catalog to include galaxies brighter than about 0.4 L^*; the number of galaxies of a given brightness placed within the simulation is determined by drawing galaxies from the SDSS galaxy luminosity function (Blanton et al. 2003), with 1.6 mag of luminosity

evolution assumed per unit redshift (the same assumption made by our cluster-finding algorithm). The choice of which particle these galaxies are assigned to is determined by relating the particle overdensities (on a mass scale of $\sim 1 \times 10^{13} \, M_\odot$) to the two-point correlation function of the particles; these particles are then chosen to reproduce the luminosity-dependent correlation function as measured in the SDSS by Zehavi et al. (2004).

Finally, colors are assigned to each galaxy by measuring their local galaxy density (here, the fifth-nearest neighbor within a redshift slice), and assigning to them the colors of a real SDSS galaxy with similar luminosity and local density. The local density measure for SDSS galaxies is taken from a volume-limited sample of the CMU-Pitt DR4 Value Added Catalog. This method produces mock galaxy catalogs that reproduce the luminosity and color correlation function of the real sky. The created mock galaxy sample therefore provides a unique tool to assess the performance of the SDSS cluster-finding algorithms in terms of completeness and purity, as well as how the observables of the detected clusters correspond to dark matter halos, assuming that galaxy clusters trace the underlying halo population in the universe.

Since precise spectroscopic redshift measurements are only available for the SDSS main sample galaxies (Strauss et al. 2002) and LRGs (Eisenstein et al. 2001), we must use photometric redshift estimates for most of the galaxies. In order to accurately reproduce this scenario in the simulations, we scatter the given redshifts of mock galaxies according to the error distributions of photometric redshift estimates, which are obtained by calibrating a sample of $\sim$140,000 SDSS photometric redshifts to their known corresponding spectroscopic measurements from the SDSS spectroscopic survey and various other sources such as CNOC2 (Yee et al. 2000), CFRS (Lilly et al. 1995a), DEEP (Weiner et al. 2005), and 2SLAQ LRG (Padmanabhan et al. 2005). The photometric redshifts were computed using a neural network technique by Lin et al. (2006) and H. Lin et al. (2008, in preparation); see also the short discussion in the SDSS DR5 data release paper (Adelman-McCarthy 2007). The comparison between calculated photometric redshifts and measured spectroscopic redshifts is shown in Figure 1 for both the red and blue galaxy samples. The
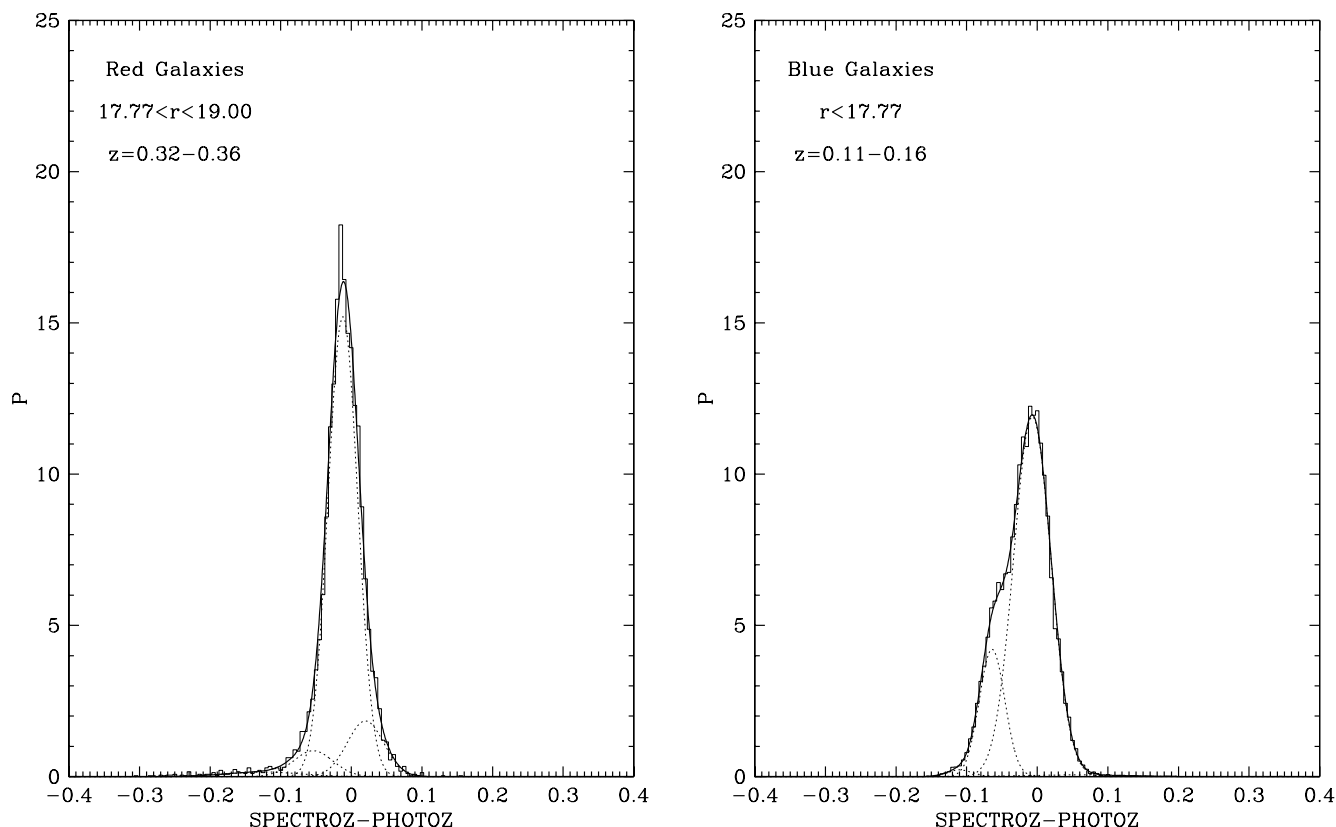
Fig. 2.—Examples of multiple Gaussian fits for the error distributions of computed photometric redshifts. The derived fitting parameters are used to scatter the known redshifts of mock galaxies in order to simulate the practice with real SDSS data.

distributions of sampled redshift uncertainties are derived for different magnitudes and redshift bins, and are found to be well described by a combination of multiple Gaussian fits, as shown, for example, in Figure 2. The resulted fitting parameters are used for the scattering of mock galaxy redshifts in the simulation. In applying the cluster-finding technique to the real SDSS data, however, we use the photo-z errors that are based on the nearest neighbor error estimate method (NNE; Lin et al. 2006) instead of deriving "empirical" error estimates collectively; this makes it possible to get an estimate of the error for each individual object, and better constrains the photometric redshift uncertainty, especially for galaxy samples with photo-z errors that depend strongly on magnitudes and the actual redshifts. We find that the computed errors correspond reasonably well with the empirical ones derived from statistics, with exceptions only for catastrophic objects. More details will be discussed in a subsequent paper on the application of the modified adaptive matched-filter technique with SDSS data.

To summarize, the implementation of a simulation of observed galaxy redshifts in a mock sample proceeds as follows: for galaxies that satisfy the SDSS spectroscopic target selection criteria, we take the given galaxy redshifts as spectroscopic measurements, while for the rest of the sample we use scattered redshifts to mimic the photometric redshift estimates. As discussed above in § 2, there are a few percent of such galaxies that fall into the redshift and magnitude ranges where no good calibrations are available. For these galaxies, we treat them as if there were no redshift information at all to put into the algorithm. We also impose on the mock galaxy catalog an apparent magnitude cut ($r < 21$), which we also intend to adopt in the SDSS imaging sample. The procedure described above thus provides the a mock catalog with the characteristics most similar to the SDSS survey, and allows

us to explore the performance of the cluster-finding algorithm on real SDSS data.

The modified matched-filter algorithm is then run on the mock galaxy catalog, and the detected clusters are compared with matched known halos given in the simulation. We find that the matches are generally robust against details of the matching techniques, as pointed out by Miller et al. (2005; although see also the discussion of various matching algorithms in Rozo et al. 2007). Here we adopt a matching criterion of projected separation between the detected cluster and the candidate matched halo within the virial radius r_{200} and with a redshift difference $\Delta z < 0.05$. To evaluate the completeness of the cluster sample, we match each dark halo to the nearest detected cluster within the projected cluster r_{200} and Δz of 0.05, while in measurement of purity, we match clusters to their corresponding halos by applying the same criteria. In the case of multiple matches, which are possible when using these matching algorithms, we designate the most massive halo within the searching space as the real match. Other methods have also been tried in efforts to refine the matching process, but no significant changes have been found in the final results.

4. RESULTS AND DISCUSSIONS

In this section, we present the results of running the modified adaptive matched-filter algorithm on the simulation-based mock catalogs. These include the completeness and purity check of the detected cluster sample, the derived cluster properties, such as estimated redshift and richness, and the expected scaling relations that would link the observed clusters to true halo distributions.

4.1. Completeness and Purity Check

We define the completeness C of the selected cluster sample as a cumulative function of M_{200}, the mass within the virial
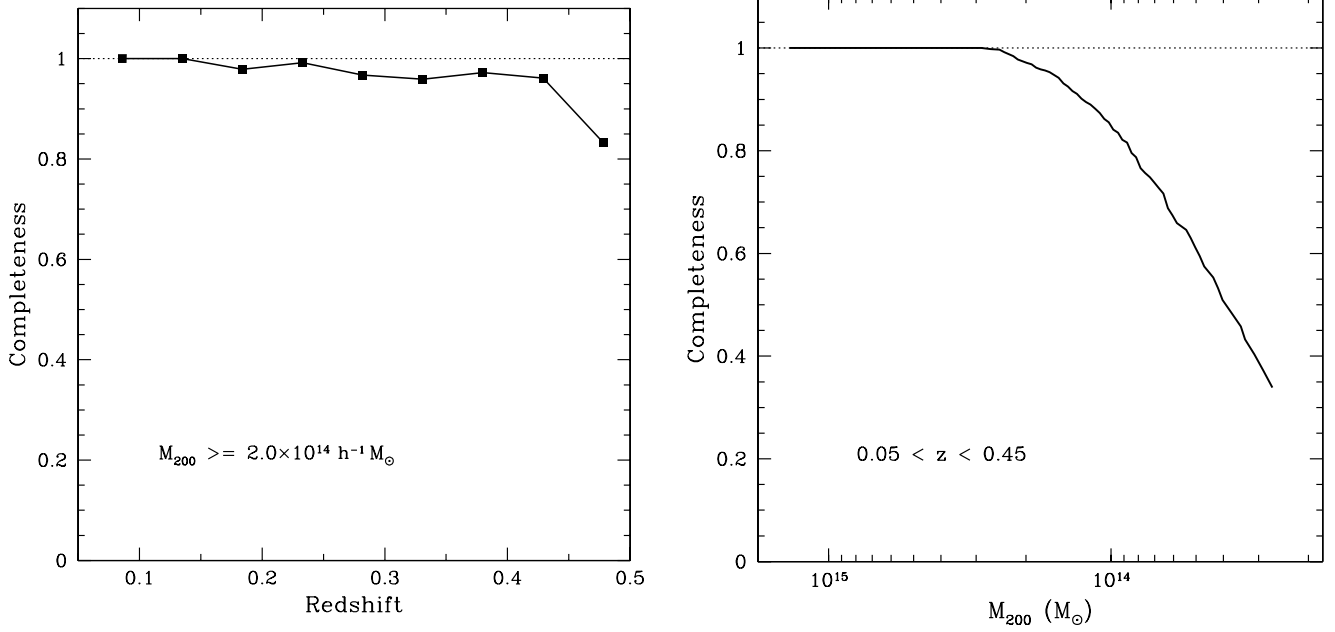
Fig. 3.—Completeness of the detected cluster sample as a function of redshift (*left*) and the virial mass of matched halos (*right*). The sample shows a consistent completeness of >95% complete for halos with $M_{200} > 2.0 \times 10^{14} \, h^{-1} \, M_\odot$, and is ∼85% complete for halos with $M_{200} > 1.0 \times 10^{14} \, h^{-1} \, M_\odot$ in the redshift range of $0.05 < z < 0.45$.

radius inside which the overdensity is 200 times the critical density:

$$C(M_{200}) = \frac{N_{\rm found}}{N_{\rm total}}, \qquad (11)$$

where $N_{\rm found}$ is the number of halos with mass greater than M_{200} matched to clusters, and $N_{\rm total}$ is the total number of halos above that mass.

Figure 3 shows the completeness of the detected cluster sample as a function of redshift (*left*) and the virial mass of matched dark matter halos (*right*). The cluster sample, which has a richness cut at $\Lambda_{200} > 20$, is over 95% complete for objects with $M_{200} > 2.0 \times 10^{14} \, h^{-1} \, M_\odot$, and ∼85% complete for objects with masses above $1.0 \times 10^{14} \, h^{-1} \, M_\odot$ in the redshift range of $0.05 < z < 0.45$. As we will find in the subsequent discussion of cluster scaling relations, the richness cut we impose on the cluster sample contributes to some of the incompleteness for less massive objects because of the large scatter in the cluster richness-mass relation; many of the matched clusters at ∼$1.0 \times 10^{14} \, h^{-1} \, M_\odot$ are simply scattered below the richness cut, and thus are not included in the computation of the completeness. This can be relieved by lowering the richness cut of the cluster sample, although we choose to retain this cut for the purity considerations below.

Figure 3*a* also shows that the completeness level of the cluster sample remains almost flat out to $z \sim 0.45$, beyond which it suffers a significant decline. This is at least partly due to the volume limit of the mock catalog, which only extends to $z = 0.58$. When we scatter the given galaxy redshifts with photometric redshift errors, which become large around $z \sim 0.5$, many of the galaxies near the far edge of the light cone are scattered away, while fewer galaxies are shifted into that range, since they are absent from the simulation. The apparent magnitude cut we have applied to the mock galaxy sample may also contribute to incompleteness at high redshift. Taking into consideration the necessary *k*-corrections, the galaxy sample is no longer complete down to the luminosity of $0.4 \, L^*$, which is the limit assumed throughout the simulation tests. The matched filter therefore loses some ability to detect less rich

systems at redshifts of $z \sim 0.5$ and beyond, since many fewer galaxies are bright enough to be observable at that distance in the current survey. We have not investigated these effects in detail, although the onset of clear incompleteness corresponds well to the distance at which they become important.

We similarly define the purity P of the selected cluster sample as a cumulative function of cluster richness Λ_{200}, which is the total cluster luminosity in units of L^* inside its virial radius r_{200}:

$$P(\Lambda_{200}) = \frac{N_{\rm match}}{N_{\rm total,\Lambda}}, \qquad (12)$$

where $N_{\rm match}$ is the number of clusters with richness greater than Λ_{200} matched to halos, and $N_{\rm total,\Lambda}$ is the total number of clusters with richness above Λ_{200}.

The results of the purity check for the obtained cluster catalog are shown in Figure 4. The sample is over 95% pure for clusters with $\Lambda_{200} > 30$, and around 90% pure for clusters with $\Lambda_{200} > 20$ over the whole redshift range out to $z \sim 0.45$. As will be shown in the richness-mass relationship below, these two richness thresholds correspond to $M_{200} \sim 6.0 \times 10^{13} \, h^{-1} \, M_\odot$ and $M_{200} \sim 4.0 \times 10^{13} \, h^{-1} \, M_\odot$, respectively. It is worth noting that the lower purity for $\Lambda_{200} > 20$ is clearly going to be affected by halo incompleteness in the simulation, since some of the matched halos for this richness will fall below the mass resolution of the halo catalog, which means that the purity we have derived above is probably a lower limit, following a logic similar to the completeness arguments.

To ensure a reasonably high purity of selected clusters, we therefore apply a $\Lambda_{200} > 20$ cut for the cluster catalog, which is used for analysis of completeness, as well as cluster-derived properties and scaling relations. The purity measurement shows a slight but notable uptrend in the last redshift bin of $z \sim 0.45-0.5$, which could be similarly explained by the arguments above in the completeness discussions. This reflects a shift in the richness-mass scaling relation at the high redshift end, where clusters with the same richness measurements may correspond to systems that are actually richer and more massive, because of the
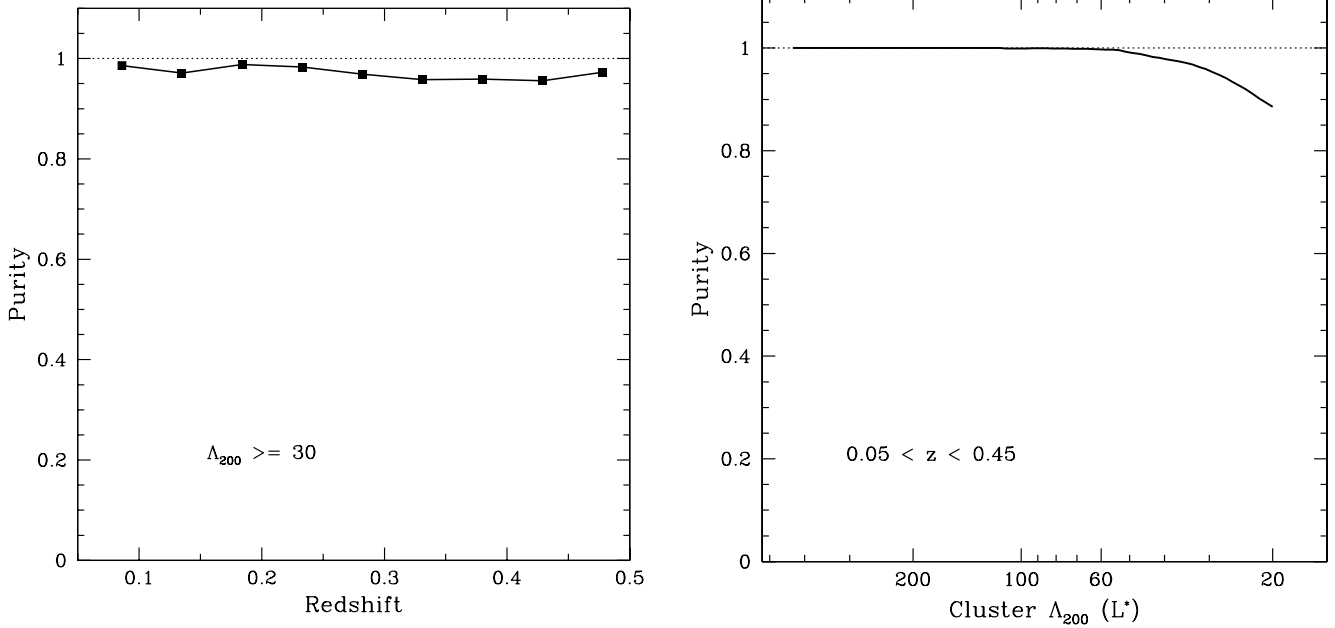
Fig. 4.—Purity of the detected cluster sample as a function of redshift (*left*) and the cluster richness (*right*). The derived catalog is over 95% pure for clusters with $\Lambda_{200} > 30$ and around 90% pure for $\Lambda_{200} > 20$ in the redshift range of $0.05 < z < 0.45$.

underrepresentation of galaxies that are observable in that redshift range. It is therefore wise to limit the current cluster catalog to a redshift of $z = 0.45$ in order to extract a uniform sample for statistical use, although the catalog using real SDSS data may well go deeper reliably.

4.2. Derived Cluster Properties and Scaling Relations

As is discussed in § 2, for each selected cluster a redshift estimate is found for the system by the matched filter that optimizes the detection likelihood at the galaxy position given as cluster center. This measurement is then taken as the estimated redshift for the cluster. Since all the halos have known redshifts in the simulation, by matching the detected clusters to halos following the procedure described in §3, we can compare the derived cluster redshifts with the true redshifts of associated halos.

Figure 5 illustrates the comparison between estimated cluster redshifts and known halo redshifts. For clusters with redshifts below $z = 0.25$ where spectroscopic redshift measurements are often available for member galaxies, the derived cluster redshift estimates precisely reproduce the true redshifts of corresponding dark halos. The inclusion of spectroscopic information of input galaxies markedly sharpens the cluster detection likelihood in the line-of-sight dimension, and thus provides accurate measurements of the cluster redshifts. In the higher redshift range, where spectroscopic measurements become rare and photometric estimates dominate, the plot illustrates a larger dispersion, while the matched filter still gives robust determinations of cluster redshifts, even with only photometric galaxy redshift information for inputs. We find that the accuracy of the redshift estimates does increase with cluster richness as expected, although this is mostly accounted for by the higher fraction of cluster galaxy members with spectroscopic measurements inside these systems. There is a slight uptrend bias seen at the redshift of $z \sim 0.45$, which we see as a similar indication of the incompleteness of the input galaxy sample near the high end of the redshift range for this mock catalog because of the volume limit and magnitude cut. The estimated cluster redshift determined from maximum likelihood tends to drift toward smaller values in some cases, since the detection probability at

higher redshift is suppressed by such effects. We also note the existence of a few serious outliers, which probably represent the occasional existence of a mismatch between relevant clusters and dark halos due to projection effects or false-positive detections.

The normalized cluster richnesses Λ_{200} are also compared with the virial mass M_{200} of matched halos. The results are shown in Figure 6. We find that the richness-mass scaling relation follows

$$\Lambda_{200} = (47.2 \pm 4.1) \times \left(\frac{M_{200}}{10^{14}\ h^{-1}\ M_{\odot}}\right)^{1.03 \pm 0.04}, \quad (13)$$

which is a roughly linear fit. Whether this is correct or not clearly depends on the details of the simulation input, and the way the
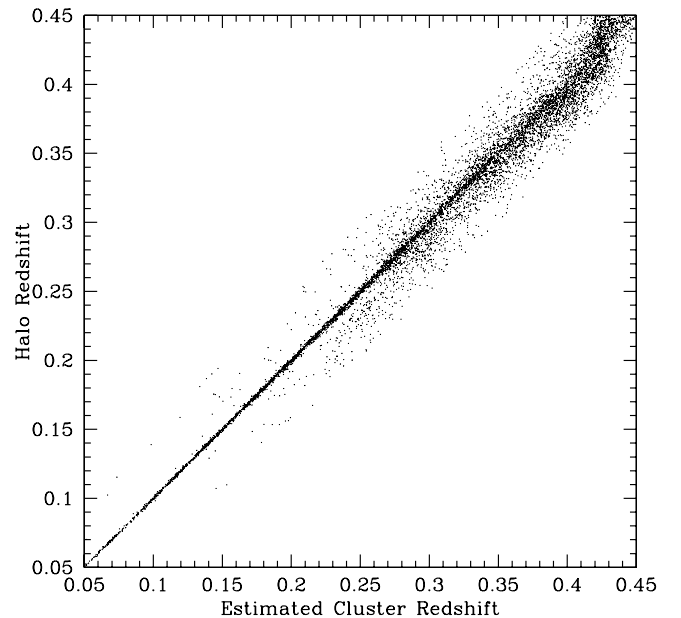

Fig. 5.—Comparison between estimated cluster redshifts and known redshifts of matched halos.
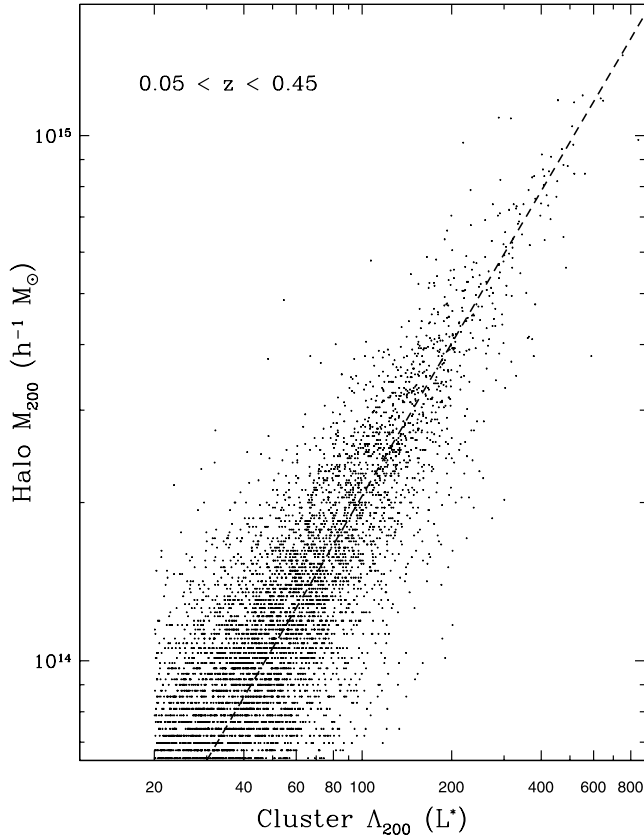
FIG. 6.—Comparison between derived cluster richness and the virial mass of matched halos. The cluster richness Λ_{200} is the total luminosity of the cluster in units of L^* inside its virial radius r_{200}.

FIG. 7.—Comparison between derived cluster richness and the total luminosity of matched halos in units of L^*. The cluster richness Λ_{200} is the total luminosity of the cluster in units of L^* inside its virial radius r_{200}.

fragmentation and merging, differing shapes between the galaxy and mass distributions, and, even further, the variable mass-to-light ratios inside the systems incorporated in the current dark matter simulations. Despite these intrinsic dispersions, the richness-mass scaling relation shows a strong linear correspondence between the

simulation was constructed gives no easy clue to what the results should be. What is important in this test, however, is that we recover what is present in the simulations, not what might or might not be present in the real universe. To that end, we have constructed three more plots. The first, Figure 7, compares the cluster richness determined by the present algorithm with the total three-dimensional luminosity of the matched halos; the agreement is very good, with no bias evident at either the sparse or the rich end. Given this agreement and the results of Figure 6, the next plot, Figure 8, of the 3D halo luminosity versus the 3D halo mass, contains no surprises. The simulated halo mass is, in fact, linear with its total luminosity, and we recover this relationship.

Figure 9 compares the derived cluster virial radius r_{200} from the cluster-finding algorithm and the r_{200} determined from 3D galaxy overdensities. The agreement is excellent at small virial radii, although there is a strong hint that the algorithm slightly overestimates large virial radii by 7% or thereabouts. This is almost certainly due to the assumption of a single NFW profile to describe the cluster; neighboring halos have rather different effects in the cylinder to which the algorithm is sensitive and the corresponding sphere in the simulations, but it is gratifying that the effects are as small as they are. These results further justify our choice to refer our richness measurements to the commonly used virial radius determined from galaxy overdensities.

It is, however, clear that the scatter in the richness-mass relation derived from the cluster finding algorithm (Fig. 6) is somewhat larger than that of the intrinsic richness-mass relation in the simulations (Fig. 8), which can be read as an indication of complications in the cluster-halo matching process, e.g., the inevitable difference between the cluster finder and the halo finder regarding
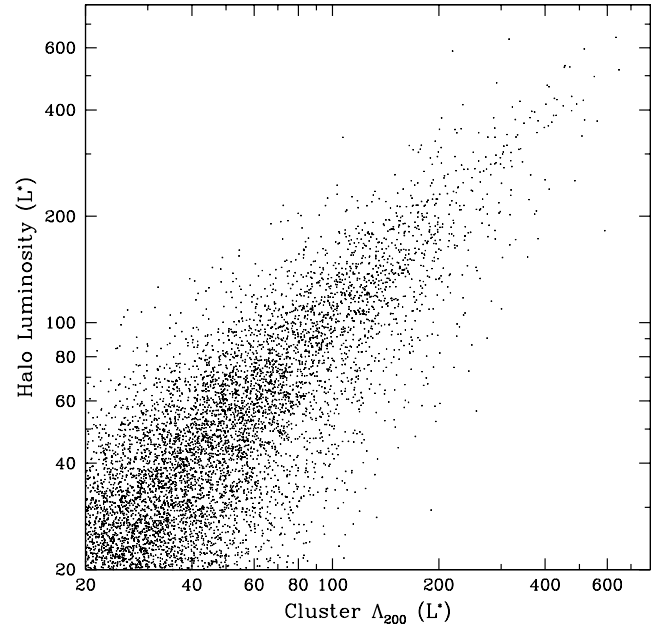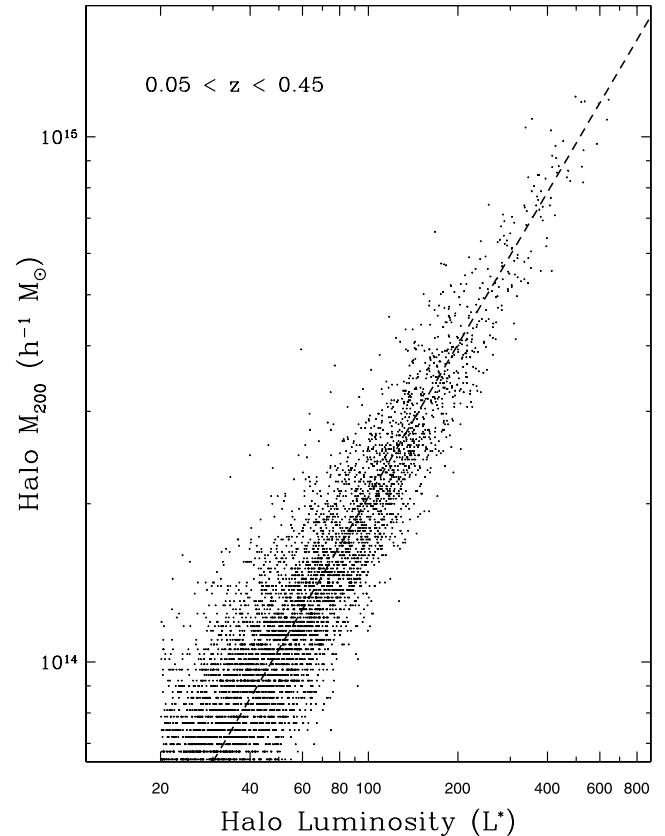

FIG. 8.—Comparison between the virial mass of matched halos and their luminosities in units of L^*. The dashed line is the best-fit cluster richness-mass scaling relation given in Fig. 6.
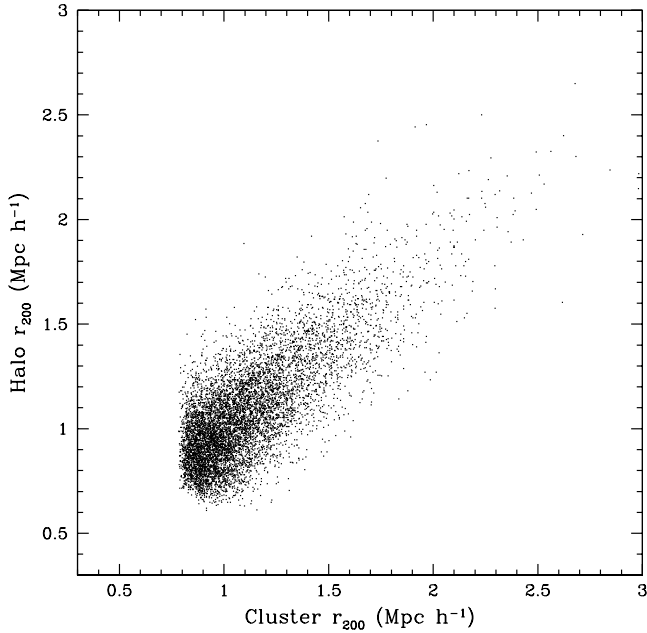
Fig. 9.—Comparison between derived cluster virial radius r_{200} and the halo r_{200}, determined by galaxy overdensities.

of cluster mass are absent here; however, the fact that we recover a linear relation from our 3D simulations indicates that we should be able to investigate a possibly more complex relationship in the real universe.

5. CONCLUSIONS

We present a modified matched-filter algorithm that is designed to construct a comprehensive cluster catalog from the Sloan Digital Sky Survey, but is applicable to any deep photometric survey. The technique is fully adaptive to 2D, 2.5D, and 3D optical surveys, as well as to various cluster scales and substructures.

The cluster-finding algorithm has been tested against a realistic mock SDSS catalog from a large N-body simulation. The results suggest that the selected cluster sample is $\sim$85% complete and over 90% pure for systems more massive than $1.0 \times 10^{14} h^{-1} M_\odot$ with redshifts up to $z = 0.45$. The estimated cluster redshifts derived from the maximum likelihood analysis show small errors with $\Delta z < 0.01$, and the normalized cluster richness measurements fit linearly with the virial mass of matched halos, the correct relation in this simulation. This offers hope that the (very likely nonlinear) relation between richness and halo mass that exists in the real universe can be investigated with these techniques.

observables and the mass, and thus makes it possible to extract the true halo distribution in the universe from the observed cluster abundance and correlation functions. It is important to note that the simulation from which the catalog was made is a dark-matter-only simulation, and thus effects which may well exist in real clusters and which can affect the baryon fraction in the intracluster gas and galaxies (see, for example, Kravtsov et al. 2005) as a function

REFERENCES

Abell, G. O. 1958, ApJS, 3, 211
Abell, G. O., Corwin, H. G., & Olowin, R. P. 1989, ApJS, 70, 1
Adelman-McCarthy, J. K. 2007, ApJS, 172, 634
Annis, J., et al. 2002, BAAS, 34, 778
Bahcall, N. A. 1988, ARA&A, 26, 631
Bahcall, N. A., et al. 2003, ApJS, 148, 243
———. 1999, Science, 284, 1481
Baldry, I. K., et al. 2005, MNRAS, 358, 441
Bartelmann, M. 1996, A&A, 313, 697
Blanton, M. R., et al. 2003, ApJ, 592, 819
Böhringer, H., et al. 2001, A&A, 369, 826
———. 2004, A&A, 425, 367
Carlstrom, J. E., Joy, M. K., Grego, L., Holder, G. P., Holzapfel, W. L., Mohr,
    J. J., Patel, S., & Reese, E. D. 2000, Phys. Scr. T, 85, 148
Christlein, D. 2000, ApJ, 545, 145
Croton, D. J., et al. 2005, MNRAS, 356, 1155
Dalton, G. B., Efstathiou, G., Maddox, S. J., & Sutherland, W. J. 1994,
    MNRAS, 269, 151
Dalton, G. B., Maddox, S. J., Sutherland, W. J., & Efstathiou, G. 1997,
    MNRAS, 289, 263
Dressler, A. 1980, ApJ, 236, 351
———. 1984, ARA&A, 22, 185
Dressler, A., & Gunn, J. E. 1992, ApJS, 78, 1
Ebeling, H., Edge, A. C., Bohringer, H., Allen, S. W., Crawford, C. S., Fabian,
    A. C., Voges, W., & Huchra, J. P. 1998, MNRAS, 301, 881
Edge, A. C., Stewart, G. C., Fabian, A. C., & Arnaud, K. A. 1990, MNRAS,
    245, 559
Eisenstein, D. J., et al. 2001, AJ, 122, 2267
Evrard, A. E. 1989, ApJ, 341, L71
Evrard, A. E., et al. 2002, ApJ, 573, 7
Folkes, S., et al. 1999, MNRAS, 308, 459
Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider,
    D. P. 1996, AJ, 111, 1748
Geller, M. J., & Huchra, J. P. 1983, ApJS, 52, 61

Gioia, I. M., Maccacaro, T., Schild, R. E., Wolter, A., Stocke, J. T., Morris,
    S. L., & Henry, J. P. 1990, ApJS, 72, 567
Gladders, M. D., & Yee, H. K. C. 2000, AJ, 120, 2148
———. 2005, ApJS, 157, 1
Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke,
    M., & Bartelmann, M. 2005, ApJ, 622, 759
Goto, T., et al. 2002, AJ, 123, 1807
Gunn, J. E., et al. 1998, AJ, 116, 3040
Hansen, S. M., McKay, T. A., Wechsler, R. H., Annis, J., Sheldon, E. S., &
    Kimball, A. 2005, ApJ, 633, 122
Henry, J. P. 2000, ApJ, 534, 565
Högbom, J. A. 1974, A&AS, 15, 417
Hogg, D. W., et al. 2003, ApJ, 585, L5
Huchra, J. P., & Geller, M. J. 1982, ApJ, 257, 423
Ilbert, O., et al. 2005, A&A, 439, 863
Kawasaki, W., Shimasaku, K., Doi, M., & Okamura, S. 1998, A&AS, 130, 567
Kepner, J., Fan, X., Bahcall, N., Gunn, J., Lupton, R., & Xu, G. 1999, ApJ,
    517, 78
Kim, R. S. J., et al. 2002, AJ, 123, 20
Kochanek, C. S., White, M., Huchra, J., Macri, L., Jarrett, T. H., Schneider,
    S. E., & Mader, J. 2003, ApJ, 585, 161
Koester, B. P., et al. 2007, ApJ, 660, 239
Kravtsov, A. V., Nagai, D., & Vikhlinin, A. A. 2005, ApJ, 625, 588
Lilly, S. J., Le Fevre, O., Crampton, D., Hammer, F., & Tresse, L. 1995a, ApJ,
    455, 50
Lilly, S. J., Tresse, L., Hammer, F., Crampton, D., & Le Fevre, O. 1995b, ApJ,
    455, 108
Lin, H., Yee, H. K. C., Carlberg, R. G., Morris, S. L., Sawicki, M., Patton,
    D. R., Wirth, G., & Shepherd, C. W. 1999, ApJ, 518, 533
Lin, H., et al. 2006, BAAS, 38, 1196
Lin, Y.-T., & Mohr, J. J. 2004, ApJ, 617, 879
———. 2007, ApJS, 170, 71
Lin, Y.-T., Mohr, J. J., & Stanford, S. A. 2004, ApJ, 610, 745
Loveday, J. 2004, MNRAS, 347, 601

Loveday, J., Peterson, B. A., Efstathiou, G., & Maddox, S. J. 1992, ApJ, 390, 338

Lumsden, S. L., Nichol, R. C., Collins, C. A., & Guzzo, L. 1992, MNRAS, 258, 1

Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., Kent, S., & Yasuda, N. 2001, in ASP Conf. Ser. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne (San Francisco: ASP), 269

Miller, C. J., et al. 2005, AJ, 130, 968

Mo, H. J., Yang, X., van den Bosch, F. C., & Jing, Y. P. 2004, MNRAS, 349, 205

Mohr, J. J., Carlstrom, J. E., & the Sza Collaboration. 2002, in ASP Conf. Ser. 257, AMiBA 2001: High-z Clusters, Missing Baryons, and CMB Polarization, ed. L.-W. Chen et al. (San Francisco: ASP), 43

Mullis, C. R., et al. 2003, ApJ, 594, 154

Nagai, D., & Kravtsov, A. V. 2005, ApJ, 618, 557

Nagamine, K., Fukugita, M., Cen, R., & Ostriker, J. P. 2001, MNRAS, 327, L10

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563

Padmanabhan, N., et al. 2005, MNRAS, 359, 237

Pierpaoli, E., Anthoine, S., Huffenberger, K., & Daubechies, I. 2005, MNRAS, 359, 261

Pierpaoli, E., Borgani, S., Scott, D., & White, M. 2003, MNRAS, 342, 163

Pierpaoli, E., Scott, D., & White, M. 2001, MNRAS, 325, 77

Postman, M., Huchra, J. P., & Geller, M. J. 1992, ApJ, 384, 404

Postman, M., Lubin, L. M., Gunn, J. E., Oke, J. B., Hoessel, J. G., Schneider, D. P., & Christensen, J. A. 1996, AJ, 111, 615

Ramella, M., Pisani, A., & Geller, M. J. 1997, AJ, 113, 483

Romer, A. K., et al. 2000, ApJS, 126, 209

Rosati, P., della Ceca, R., Norman, C., & Giacconi, R. 1998, ApJ, 492, L21

Rozo, E., Wechsler, R. H., Koester, B. P., Evrard, A. E., & McKay, T. A. 2007, ApJ, submitted (astro-ph/0703574)

Scharf, C., Donahue, M., Voit, G. M., Rosati, P., & Postman, M. 2000, ApJ, 528, L73

Schechter, P. 1976, ApJ, 203, 297

Schneider, P. 1996, MNRAS, 283, 837

Schuecker, P., & Boehringer, H. 1998, A&A, 339, 315

Schwarz, U. J. 1978, A&A, 65, 345

Shectman, S. A. 1985, ApJS, 57, 77

Sheldon, E. S., et al. 2004, AJ, 127, 2544

Smith, J. A., et al. 2002, AJ, 123, 2121

Strateva, I., et al. 2001, AJ, 122, 1861

Strauss, M. A., et al. 2002, AJ, 124, 1810

Sutherland, W. 1988, MNRAS, 234, 159

Tremaine, S. D., & Richstone, D. O. 1977, ApJ, 212, 311

Wechsler, R. H. 2004, in Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution, ed. J. S. Mulchaey, A. Dressler, & A. Oemler (New York: Cambridge Univ. Press), 53

Weiner, B. J., et al. 2005, ApJ, 620, 595

White, M., & Kochanek, C. S. 2002, ApJ, 574, 24

Wittman, D., Tyson, J. A., Margoniner, V. E., Cohen, J. G., & Dell'Antonio, I. P. 2001, ApJ, 557, L89

Yee, H. K. C., et al. 2000, ApJS, 129, 475

York, D. G., et al. 2000, AJ, 120, 1579

Zehavi, I., et al. 2004, ApJ, 608, 16

Zheng, Z., et al. 2005, ApJ, 633, 791

Zwicky, F., Herzog, E., & Wild, P. 1968, Catalogue of Galaxies and of Clusters of Galaxies (Pasadena: Caltech)